(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42327 A2

(51) International Patent Classification[7]: C07K 14/415

(21) International Application Number: PCT/EP01/13536

(22) International Filing Date: 21 November 2001 (21.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00125523.1 21 November 2000 (21.11.2000) EP

(71) Applicant (*for all designated States except US*): **SOCIETE DES PRODUITS NESTLE S.A.** [CH/CH]; P.O. Box 353, CH-1800 Vevey (CH).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **KOCHHAR, Sunil** [IN/CH]; Chemin du Grammont 2, CH-1073 Savigny (CH). **HANSEN, Carl, Erik** [NO/CH]; Chemin du Bois-Murat 20, CH-1066 Epalinges (CH). **JUILLERAT, Marcel, Alexandre** [CH/CH]; Chemin Pra-Dom-Nicod 10, Vers-Che-Les-Blancs, CH-1000 Lausanne 26 (CH). **Mc CARTHY, James** [US/FR]; 298, Vallée de Vautruchot, F-37210 Noizay (FR).

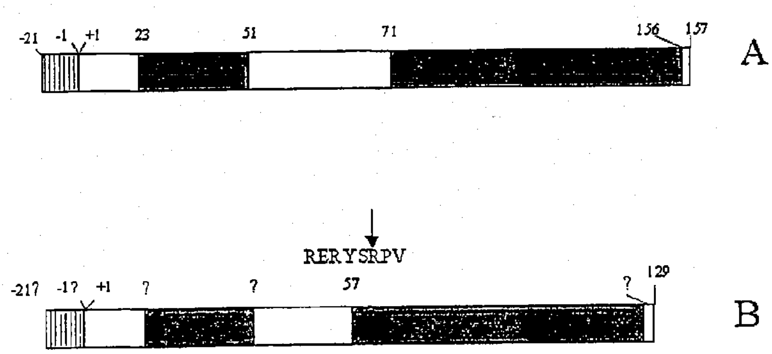
(74) Agent: **STRAUS, Alexander**; Becker, Kurig, Straus, Bavariastrasse 7, 80336 Munich (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: A NOVEL COCOA ALBUMIN AND ITS USE IN THE PRODUCTION OF COCOA AND CHOCOLATE FLAVOUR




```
1   RPVSKHLDSC CQQLEKLDTP CRCPGLKQAV QQQAEEGEFG REELQEMYET
51  VDKIMNKCDV EPGRCNLQPR NWF (SEQ ID No 4)
```

Schematic representation of Brassica 2S protein processing (A) and the predicted processing sites for the 2S protein of *T. cacao* (B) as well as the amino acid sequence of the mature polypeptide.

(57) Abstract: A novel 2S cocoa albumin was isolated, purified and identified from cocoa beans. Enymatic hydrolysis of the protein generated a pool of flavour precursors, peptides and amin o acids that resultd in formation fo cocoa flavour upon heating with sugars. The DNA encoding a precursor cocoa 2S protein was isoltated from immature Theobroma cocoa seeds.

WO 02/42327 A2

Published:
— *without international search report and to be republished upon receipt of that report*

A novel cocoa albumin and its use in the production of cocoa and chocolate flavour

The present invention relates to a novel cocoa polypeptide and the DNA encoding it. In particular, the present invention pertains to the use of said polypeptide and/or fragments thereof in the production of cocoa/chocolate flavour.

In processing cocoa beans the generation of the typical cocoa flavour requires two steps, the fermentation step and the roasting step. During fermentation the pulp surrounding the beans is degraded by micro-organisms with the sugars contained in the pulp being essentially transformed to acids. Fermentation also results in a release of peptides exhibiting differing sizes and a generation of a high level of free hydrophobic amino acids. This latter finding led to the hypothesis that proteolysis occurring during fermentation is not due to a random protein hydrolysis but seems to be rather based on the activity of specific endoproteinases. This specific mixture of peptides and hydrophobic amino acids is deemed to represent cocoa-specific flavour precursors. During the second step of cocoa flavour production, the roasting step, the oligopeptides and amino acids generated at the stage of fermentation obviously undergo a Maillard reaction with reducing sugars present eventually yielding the substances responsible for the cocoa flavour as such.

So far, research has tried to uncover the molecular pathway of producing cocoa flavour precursors in characterizing enzymes involved in said process and the relevant polypeptide(s), from which the peptides and/or free amino acids are produced.

As for the enzymes many different endo- and exoproteinase activities have been found to participate in the production of cocoa flavour precursors, such as an aspartic endoproteinase activity (Voigt et al., J. Plant Physiol. **145** (1995), 299-307), which accumulates with the vicilin-class (7S) globulin during bean ripening or a cysteine endoproteinase activity, which increases during the germination process when degradation of globular storage protein

increases during the germination process when degradation of globular storage protein occurs (Biehl et al., Cocoa Research Conference, Salvador, Bahia, Brasil, 17-23 Nov. 1996). Moreover, a carboxypeptidase activity has been identified which preferentially splits hydrophobic amino acids from the carboxy-terminus of peptides.

Apart from the enzymes also the protein source of the peptides/amino acids seems to be of importance for the generation of cocoa flavour precursors.

During cocoa bean fermentation the percentage reduction of protein concentration observed for vicilin and albumin was 88.8% and 47.4%, respectively (Amin et al. J. Sci. Food Agric. **76** (1998), 123-128). When peptides obtained by proteolysis of the globulin (vicilin) fraction were post-treated with carboxypeptidase, preferentially hydrophobic amino acids were released and a typical cocoa aroma was detected after roasting in the presence of reducing sugars (Voigt et al., Food Chem. **50** (1994), 177-184). Contrary to that, the predominant amino acids released from the albumin-derived peptides were aspartic acid, glutamic acid and asparagine and no cocoa aroma could be detected. It was therefore concluded that cocoa-specific aroma precursors are mainly derived from the vicilin-like globulins of cocoa, which constitute more than 30 % of the total protein contents in the mature cocoa seed. Consequently, the mixture of hydrophobic free amino acids and remaining oligopeptides required for the generation of the typical cocoa flavour components seem to be determined by the particular structure of the cocoa vicilin-class globulins.

Although it is known that hydrophobic amino acids are important cocoa flavour precursors, the specific peptides responsible for generating cocoa flavour during roasting remain by and large un-characterized. Consequently, there is a need in the art to provide further structural data of such peptides and of the way they are produced from their original proteinaceous source in order to be capable to use those peptides for the production of a well-balanced cocoa and/or chocolate flavour.

In WO 91/19801 two major cocoa seed storage proteins and the DNA sequences encoding said polypeptides are disclosed. These two proteins exhibit a molecular weight of 47 kDa

and 31 kDa, respectively, and seem to be the vicilin polypeptides, which are deemed to be the source of the flavouring peptides creating the characteristic cocoa flavour. Though the polypeptides have recombinantly been provided as such no specific data as to the synthesis of the flavouring peptides have been provided.

Consequently, an object of the present invention resides in further elucidating the generation of cocoa flavour from the relevant protein material contained in cocoa.

In order to solve the above problem research has naturally focused on the vicilin polypeptides in cocoa beans, since other protein material contained therein was not considered to contribute to the generation of cocoa flavour as such. In contrast to this general belief the present inventors have now surprisingly found that a polypeptide, being a member of the albumin family, also contributes to the characteristic cocoa flavour during fermentation and roasting.

Thus, the present invention provides a novel polypeptide as identified by SEQ ID NO 1 or fragments thereof having a N-terminus comprising the amino acid sequences as identified by SEQ ID NO 2 or 3, and/or heterodimers of said fragments. In a preferred embodiment the mature polypeptide as identified by SEQ ID No 4 is provided.

According to another aspect the present invention provides a nucleic acid as identified by SEQ ID NO 5, or a derivative thereof, encoding any of the above polypeptide(s). The present nucleic acids also comprise DNA molecules that are derived from the nucleic acid identified by SEQ ID NO 5 by the degeneracy of the genetic code or by substituting one or more bases with the proviso that a polypeptide identified by SEQ ID NO 1 will be obtained. The present invention also contemplates allelic variations of the nucleic acid indicated.

In the Figures,

Fig. 1 an SDS-PAGE analysis of different extracts of cocoa acetone powder;

Fig. 2 shows an SDS-PAGE analysis of the purified 2S albumin;

Fig. 3 shows the predicted α-helical regions and the hydrophobicity plot for the T. cacao 2S precursor protein;

Fig. 4 is a schematic representation of Brassica 2 S protein processing and the predicted processing sites for the 2S protein of T. cacao and shows the amino acid sequence of the mature polypeptide (SEQ ID NO 4) according to tryptic peptides mass fingerprints of the purified protein;

Fig. 5 shows a cocoa flavour evaluation of enzymatically hydrolyzed cocoa polypeptides;

During the studies leading to the present invention the inventors originally tried to find peptides derived from the vicilin like globulins present in cocoa. To this end, several experiments were carried out on cocoa acetone powder, wherein the 21 kDa albumin polypeptide was selectively removed. After several purification steps a substantially homogeneous protein preparation was found that showed a major band at about 9 kDa and a weak band at about 4 kDa. The protein thus obtained was sequenced and two amino acid sequences were obtained:

RPVSK HLDSC CQQLE KLDTP PRRPG LKQAV QQCA; (SEQ. ID. No. 2)

and

SKEXS CKXI (SEQ. ID. No. 3)

On the basis of these information a cDNA library prepared from T. cacao was screened for nucleic acids encoding such (a) protein(s) and a polypeptide with a theoretical molecular weight of about 17 kDa could be located. The nucleic acid and the deduced amino acid sequence is shown in SEQ ID NOs 4 and 1.

As may be seen from a comparison of the amino acid sequences obtained from sequencing the purified protein and the amino acid sequence (open reading frame) derived from the nucleic acid both of the sequences are contained in the open reading frame of the subject

nucleic acid molecule indicating a post-translational processing of a precursor molecule as represented by the 17 kDa polypeptide.

As is known from other species, e.g. rape seed, a precursor polypeptide of the 2S protein is subjected to different post-translational processing steps including the generation of two subunits that are held together by intra- and inter-chain disulfide bonds. These two peptides are produced by removal of peptides at the N-terminus, between the subunits and at the C-terminus of the precursor molecule. On the basis of the information provided a similar mechanism seems to take place with the 2 S-polypeptide of T. cacao, as evidenced by the occurrence of two different N-terminal sequences.

Consequently, according to a preferred embodiment the present invention provides a polypeptide, which is derived from the 17 kDa polypeptide, and which has a N-terminus comprising the amino acid sequence as identified under SEQ ID NO 2. This part represents a subunit of the mature 2S-polypeptide.

According to yet another preferred embodiment the present invention relates to a polypeptide derived from the 17 kDa polypeptide as described herein, and which a N-terminus comprising the amino acid sequence as identified under SEQ ID NO 3. This part of the 17 kDa polypeptide represents another subunit of the mature 2S protein.

Further, it could now be shown that the present 2S protein of T. cacao also yields peptides, which upon reaction with reducing sugars results in cocoa flour products (see below). Therefore, the present 17 kDa polypeptide or fragment thereof, preferably fragments comprising the amino acid sequence of the subunits of the mature 2S-polypeptide may be recombinantly produced to obtain cocoa flavour precursors which may be used for producing cocoa flavour.

For expression a nucleic acid coding for any of the polypeptides of the present invention may be incorporated in a suitable vector, with which a cell of interest is transformed. Since the polypeptide does not seem to be glycoslated expression in prokaryotic cells is also

possible. To this end, the nucleic acid as identified by SEQ ID NO 5 or a fragment thereof may be incorporated in an expression vector, such as pUC, pNZ124 (Platteuw et al., (1994) Appl. Env. Microbiol. **60**, 587), pGK12 (Walke et al., (1996) FEMS Microbiol. **138**, 233,) or pG+host9 (Maguin et al., (1996) J. Bacteriol **178**, 931) etc.. For expression in e.g. methylotrophic yeast *Pichia pastoris*, the vector pPICZaA as described Manual of the EasySelect™ *Pichia* Expression kit, version B (Invitrogen, The Netherlands) can be used. Heterologous expression in *Yarrowia lipolytica* can be obtained with the vector pINA1294 containing a defective *ura*3 gene (*ura*3d4) that allows direct selection for multicopy integrants (Madzak, C., Treton, B. and Blanchin-Roland, S. (2000) Strong Hybrid Promoters and integrative expression/selection vectors for Quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. J. Microbiol. Biotechnol. 2(2): 207-216) . For *Hansenula polymorpha*, B14-derived expression vectors containing the *FMD* promoter and *MOX* terminator as described in Mayer, A.F., Hellmuth, K., Schlieker, H., Lopez-Ulibarri, R., Oertel, S., Dahlems, U., Strasser, A.W.M., van Loon, A.P.G.M. (1999) Biotechnol. Bioeng. 63:373. It will be appreciated that the skilled person is well aware of arranging the corresponding nucleic acid such that an open reading frame is present, such as is e.g. necessary for producing the fragments of the 17 kDa precursor. To this end, a start codon may be positioned directly in front of the respective N-terminus of a fragment or may be positioned such that it is spaced from the polypeptide to be expressed by a linker which may support the isolation of the resulting polypeptide. Methods for introducing a nucleic acid into a vector and for transforming cells with the vector are known to the skilled person and may be found in "Maniatis and Sambrook, A Laboratory Manual, Cold Spring Harbor (1992), USA".

The cell of interest may be any cell or cell line with which the present polypeptide or a fragment thereof may be expressed. Expression in E. coli may be advantageous due to its easy handling and the option to choose a variety of different expression vectors for expressing polypeptides within the cell or in secreted form. However, the nucleic acids of the present invention may well be incorporated in cells of higher origin, such as plant cells, in particular in cells of T. cacao. In this respect over-expression of the 2S polypeptide may be achieved in a recombinant T. cacao plant by incorporating a nucleic acid as identified by

SEQ ID NO 5 into a cocoa cell using vectors suitable for plants, such as the Ti-plasmid or using the technique of homologous recombination. Such a plant will eventually yield a higher content of cocoa flavour precursors.

In the case of producing the polypeptides of the present invention by recombinant means in bacteria, yeast or in cell culture in general the polypeptide may be isolated by methods known per se and the purified 2S polypeptide may be subjected to a proteolytic degradation, using the different enzymes known to participate in the generation of cocoa flavour. In a subsequent step the flavour precursors thus obtained may be contacted with sugars such that a Maillard reaction may take place eventually obtaining cocoa/chocolate flavour.

However, substances yielding cocoa flavour are known to also beneficially affect phsiological and/or medical conditions, and may thus be used in the treatment of hypertension or mood depression. Also immune modulatory activities are known, such as improving an individuals capability to cope with bacterial challenges. The present invention therefore also envisages such usages.

The following examples illustrate the invention in a more detailed manner. It is, however, understood that the present invention is not limited to the examples but is rather embraced by the scope of the appended claims.

Example 1

Identification of a cocoa 2S albumin

Cocoa pods were obtained from experimental farms in Ecuador, Ivory Coast and Malaysia and unless stated otherwise all studies were carried out using West African Amelonado cocoa beans. Due to the high fat and polyphenol contents, proteins were extracted from Cocoa Acetone Powder (CAP). The CAP was prepared from non-defatted cocoa beans as follows: Sun-dried unfermented cocoa beans were passed through a bean crusher followed by a winnower to remove shells. The cocoa nibs were milled and the nib powder was passed through 0.8-mm sieve. The cocoa nib powder was suspended in 80 % (v/v) aqueous acetone,

stirred and subsequently centrifuged. The residue was extracted 5-times with 80 % (v/v) aqueous acetone and 3-times with 100 % acetone. The acetone powder was dried under reduced pressure.

CAP (5g) was suspended in 50 ml ice-cold sodium-acetate buffer (50 mM, pH 4.0) containing 0.1 mM pepstatin. The suspension was sonicated for 2 x 30 sec with a 10 min layover interval on ice. The suspension was centrifuged at 20,000 g for 15 min at 4 °C. The residue was extracted twice with buffer pH 4.0 (supra) followed by water extraction employing sonication. The residue from the water extract was finally extracted with 100 mM Tris-hydrochloride, pH 8.5. The supernatant was passed through a sterile 0.22 µm filter and stored at –20 °C.

SDS-PAGE analysis of pH 8.5 extracts following two exhaustive washing of the residue with pH 4 buffer followed by a water wash showed complete absence of the high intensity 21 kDa protein (Fig.1). Separation was carried out in a conventional manner (Lämmli (1970)), employing 12.5 % gels. Lane A contains low range molecular markers (Bio-Rad), lane B contains total CAP extract with 1 % SDS, lane C contains CAP extracted 2 times with 100 mM acetate buffer, pH 4,0 (supra), lane D contains a subsequent water extract of the residue of lane C and lane D contains an extract of the residue of D with 100 mM Tris-hydrochloride buffer, pH 8.5.

It could be shown that the 21 kDa protein could be essentially removed. Further under these conditions a protein having a molecular weight of about 9 kDa was detected which was further purified.

Chromatographic steps were performed at room temperature using a BioCad 20 chromatography station (Perseptive Biosystems). The frozen CAP extract was thawed overnight at 4 °C and, if necessary, adjusted to pH 8.5 with 1M Tris-chloride buffer, pH 8.5. The clear CAP extract was applied to a Resource Q column (Amersham-Pharmacia biotech) equilibrated with buffer A (50 mM Tris-bis-propane chloride, pH 8.5) at a flow rate of 5 ml/min. The column was washed with buffer A until A_{280} decreased to below 0.05. The

column was eluted with a linear gradient (20 column volumes) of NaCl 0-500 mM in buffer A. Fractions were analysed by SDS-PAGE, and those showing the 9 kDa protein were pooled, concentrated by ultra-filtration (PM-10 membrane, Amicon). The concentrated 9 kDa protein fraction was injected onto a HiLoad Superdex 30 column (26 x 600 mm) equilibrated with 50 mM sodium phosphate buffer, pH 7 containing 100 mM NaCl. The column was eluted with the same buffer and fractions were collected. The fractions showing the 9 kDa protein were pooled and concentrated by ultra-filtration. The purified protein solution was passed through a fast-desalting PD-10 column (Amersham-Pharmacia Biotech) for buffer exchange to water, sterile filtered, and stored at -20 °C.

Two successive chromatography steps, anion exchange and gel filtration resulted in apparent homogeneity of the protein preparation as judged by SDS/PAGE followed by Coomassie Brilliant Blue staining (Fig. 2). 10 µl samples were diluted 3 x with sample buffer in the presence and absence of ß-mercaptoethanol and centrifuged and electrophoresed at 100 V. The gels were stained for small peptides. Lane A contains a molecular weight marker, lane B contains a CAP extract (1 % SDS/50 mM phosphate buffer, pH 7,0) following exhaustive extraction at pH 4; lane C purified 2S albumin in the absence of ß-mercaptoethanol, and lane D purifed 2S albumin under reducing conditions.

The glycosylation was assessed employing the glycoprotein detection kit from Bio-Rad. The purified cocoa albumin was not found to be glycosylated.

The polypeptide was subjected to Edman degradation resulting in 2 amino acid sequences to be obtained, i.e. RPVSK HLDSC CQQLE KLDTP PRRPG LKQAV QQCA and SKEXS CKXI.

Example 2

Cloning of the 2 S albumin gene

Total RNA was isolated from mature and less mature seeds according to methods known per se (Maniatis, supra). Poly A+ RNA was prepared from the total cocoa seed RNA using the Oligtex kit from Qiagen following the kit instructions for 250-500 µg total RNA. The final

pellet was resuspended in 10 μL of RNase free water, and the concentration of RNA present was estimated to be approximately 5-10 ng/μL using Clontech nucleic acid "Quick Sticks".

The synthesis of cDNA from the polyA+ mRNA was carried out using a SMART PCR cDNA synthesis kit from Clontech. The method used was as described in the kit instructions. For the first strand cDNA synthesis step, 4 μL (20-40 ng) of poly A+ mRNA was used and 200 units of Gibco BRL Superscript II MMLV reverse transcriptase. The PCR step of the SMART protocol was also set up as directed in the kit instructions, except only 2 μL of the first strand reaction was added. First, 18 cycles of a PCR were run, then, 35 μL was taken out of the total reaction (100 μL) and this part of the reaction was run for a further 5 cycles of PCR.

A pool of two PCR reactions was then prepared; 40 μL of the 18 cycle PCR reaction and 15 μL of the 23 cycle PCR reaction. 2.5 μL protease K (Boehringer Mannheim, nuclease free, 14 μg/μL) was added to this cDNA mixture and the reaction was carried out at 45 °C for one hour. After a brief spin, the reaction was stopped by heating the mixture to 90 °C for 8 min. The mix was then chilled on ice, and 5 μL of T4 DNA polymerase (New England Biolabs) was added (3 units/μL), and the reaction was incubated at 14-16 °C for 30 min. Then, 25 μL of Milli Q water, 25 μL phenol (Aqua phenol), and 25 μL chloroform/isoamyl alcohol (Ready Red) was added. This mixture was vortexed, spun, and the top aqueous layer was taken. The phenol layer was re-extracted with 50 μL of water. The two resulting aqueous layers were then pooled and re-extracted with chloroform/isoamyl alcohol (Ready Red). The DNA in the aqueous layer obtained was ethanol precipitated as described above. The dried DNA obtained was resuspended in TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) and its concentration was calculated to be approximately 75 ng/μL using nucleic acid "Quick Sticks" strips from Clontech.

The cDNA was then ligated into the PCR-Script Amp SK(+) cloning vector of Stratagene. Two μL of the ligated DNA was transformed into Stratagene Ultracompetent cells XL-2 Blue as described in the instruction manual for these cells.

Eighteen randomly chosen inserts containing clones of the cDNA library were subjected to a single DNA sequencing run using the T3 primer present in the pPRC-Script Amp vector. Potential protein coding sequences of these DNA sequences were identified using the "Lasergene" suite of DNA analysis programs from DNASTAR Inc. The amino acid sequences obtained for open reading frames were then compared to the sequences obtained in the Edman degradation (example 1, above). For the 18 clones analyzed 3 clones were found to contain the same cDNA sequence encoding a protein harboring the amino acid sequences as identified for the polypeptide searched for.

The DNA insert is 718 base pairs in length and an analysis of the protein encoded by this cDNA shows that the 2S protein is probably produced first as a precursor having 150 amino acids with a calculated molecular weight of 17,125 Da and a pI of 6.15. The amino acid composition profile for the precursor 2S protein shows that the cocoa 2S protein has a relatively high level of sulfur containing amino acids.

Example 3

Biochemical characterisation of the 2S-protein

LC/ESI-MS Analysis Data of a cocoa 2S albumin:

LC/ESI-MS analysis showed the molecular weight of the mature protein to be 8513±2 Da (Fig. 5). Reduction and S-pyridinylethylation resulted in a positive shift of 630 mass units (Mr 9,145) indicating the presence of 6 cysteine residues.

Tryptic peptide mass fingerprinting:

The primary structure of the mature cocoa 2S albumin having a molecular mass of 8513±2 Da was determined by generating the tryptic peptide mass fingerprints of the reduced and pyridinylethylated albumin by RP-HPLC/ESI-MS. A total of 10 peptide masses were detected (Table 1).

Table 1
Tryptic Peptide Analysis of Cocoa 2S Albumin by LC/ESI-MS

Theoretical average $[M+H]^+$	Sequence position	Tryptic peptides	Average observed $[M+H]^+$
3829.042	4-39	ND	
1576.735	105-118	T8	1576.7
1513.673	119-130	T7	1513.6
1439.568	55-65	ND	
1303.577	83-93	T9	1513.5
836.462	76-82	ND	
775.340	135-141	T3	880.4
730.366	142-147	T5	
724.322	66-73	ND	809.4
704.340	94-99	T6	
681.287	40-45	ND	
665.362	50-54	ND	622.3
517.280	100-104	T4	505.3
505.280	131-134	T2	
466.208	148-150	ND	
388.255	46-48	ND	
349.190	1-3	ND	
304.162	74-75	T1	586.4
147.113	49-49	T10	1536.5/1447.9

ND, not detected

The mature 2S protein of plants such as Brassica napus (rape seed) and pumpkin (Hara-Nishimura et al., "Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons", Plant Physiol. 85 (1987) 440-445) are known to be post-translationally processed to generate two subunits. A comparison of the observed tryptic peptide masses of the mature protein against the translated amino acid sequence showed a 100 % amino acid sequence match to the residue 78 to 147 (SEQ ID NO 1). The peptide fragments containing the cysteine residues showed the expected positive mass shift of 105 due to S-pyridinylethylation. Every identified peptide mass was subjected to MS/MS analysis to determine either a complete or partial amino acid sequence to confirm its mapping to the amino acid sequence of the albumin. The C-terminal peptide NWF could not be detected. Also N-terminal peptides (sequence residues 1-77) could not be detetcted indicating that the 2S cocoa albumin is post-translationally processed to yield a much smaller polypeptide from its N-terminal end.

Hydrophobicity:

Analysis of an hydrophobicity plot for the cocoa 2S precursor protein (SEQ ID NO 1)

clearly indicates that the N-terminal region of this 2S protein encodes a distinct short hydrophobic region that is considered to represent the signal peptide sequence. The predicted α-helical regions for the T. cacao 2S precursor shows that the position of the N-terminal residue of large cocoa 2S fragment mapped by N-terminal sequencing (position 77 in SEQ ID NO 1) has a noticeable absence of α-helical forming sequences.

Example 4

Flavour Potential of the 2S polypeptide

Isolated cocoa polypeptide fractions (lyophilized powder), namely 21 kDa albumin (SA), 8.5 kDa albumin (2SA), insoluble vicilin protein fraction (InsV) and total polypeptide fraction (CPF) were suspended in 100 mM acetate buffer, pH 5 and digested with 1 % (w/w total protein) Flavourzyme for 16-24 h. Alternatively, the polypeptides were digested in 100 mM acetic acid, pH 3 with 1 % (w/w protein) porcine pepsin for 16-24 h. Both samples were freeze dried. A subset (at least 70 %) of pepsin hydrolyzed sample was further digested with 200 units of carboxypeptidase A. Following analytical analysis (free and total amino groups and amino acids), an identical amount of each hydrolysate was reacted with reducing sugars as described under the following section.

The process reaction flavours using amino acid residues or protein hydrolysates were prepared as follows: The reference model reaction was prepared by reacting 0.8 % Leu, 1.45 % Phe, 0.8 % Val, 1.5 % fructose, 1.5 % water (4 drops of 50 % (w/v) NaOH in 20 ml water) and 94 % propylene glycol at 125 °C (temperature of oil bath) for 60 min under reflux. The cocoa protein hydrolysate-based reaction flavours were prepared by replacing the amino acids with 1 % (w/w) of lyophilized hydrolysate. At the end of the reaction, each mixture was cooled to room temperature, and its final pH as well as optical density at 420 nm was measured. The reactants were transferred in a dark-brown bottle and stored at 15 °C until sensory profiling.

A panel of 8 persons was used to evaluate the flavour (aroma and taste) of the process

reaction on a scale of 1-10 for different flavour attributes. Tasting was performed on 0.1% (w/w) solutions in 1% (w/w) sucrose. For each sensory session an average of score data was used to evaluate the flavour potential of the various polypeptide fractions.

The results are summarized in Fig. 5, which shows the evaluation of various precursor pools generated from the enzymatically hydrolyzed cocoa polypeptide fractions in the flavour assay system. As expected the most cocoa flavour is produced by the vicilin storage protein fraction. Surprisingly, also the newly identified 2S albumin showed respectable cocoa flavour when hydrolyzed by Flavourzyme or pepsin/carboxypeptidase combination. Selection of the enzyme cocktail for extensive hydrolysis showed no remarkable difference suggesting that cocoa polypeptides harbor innate amino acid sequences for generation of cocoa flavour. The flavour quality and intensity of 2S albumin was surprisingly superior to the highly abundant 21 kDa cocoa albumin. These data strongly support the notion that 2S polypeptide together with vicilin storage protein contributes significantly to the accumulation of the potential cocoa flavour precursors.

Claims

1. A polypeptide as identified by SEQ ID. No. 1.

2. A polypeptide, derived from the polypeptide according to claim 1, having a N-terminus comprising the sequence as identified by SEQ ID. NO. 2.

3. A polypeptide, derived from the polypeptide according to claim 1, having a N-terminus comprising the sequence as identified by SEQ ID. NO. 3.

4. A polypeptide being a heterodimer consisting of a polypeptide according to claim 2 and 3.

5. A nucleic acid as identified by SEQ ID. No. 4 or a fragment thereof encoding a polypeptide according to any of the claims 1 to 3.

6. An expression vector containing one or more of the nucleic acids according to claim 5.

7. A cell containing a recombinant nucleic acid according to claim 6 or a vector according to claim 6.

8. Use of a polypeptide according to any of the claims 1 to 4, for the preparation of cocoa/chocolate flavour.

9. Use of a polypeptide according to any of the claims 1 to 4 for the preparation of a composition for the treatment of hypertension, mood depression, bacterial infections and a weakened immune condition.

10. A method for producing cocoa flavour comprising hydrolysing a polypeptide according to any of the claims 1 to 3 by an endopeptidase or in combination with exopeptidases and subjecting the resulting peptides to a reaction involving reducing sugars.

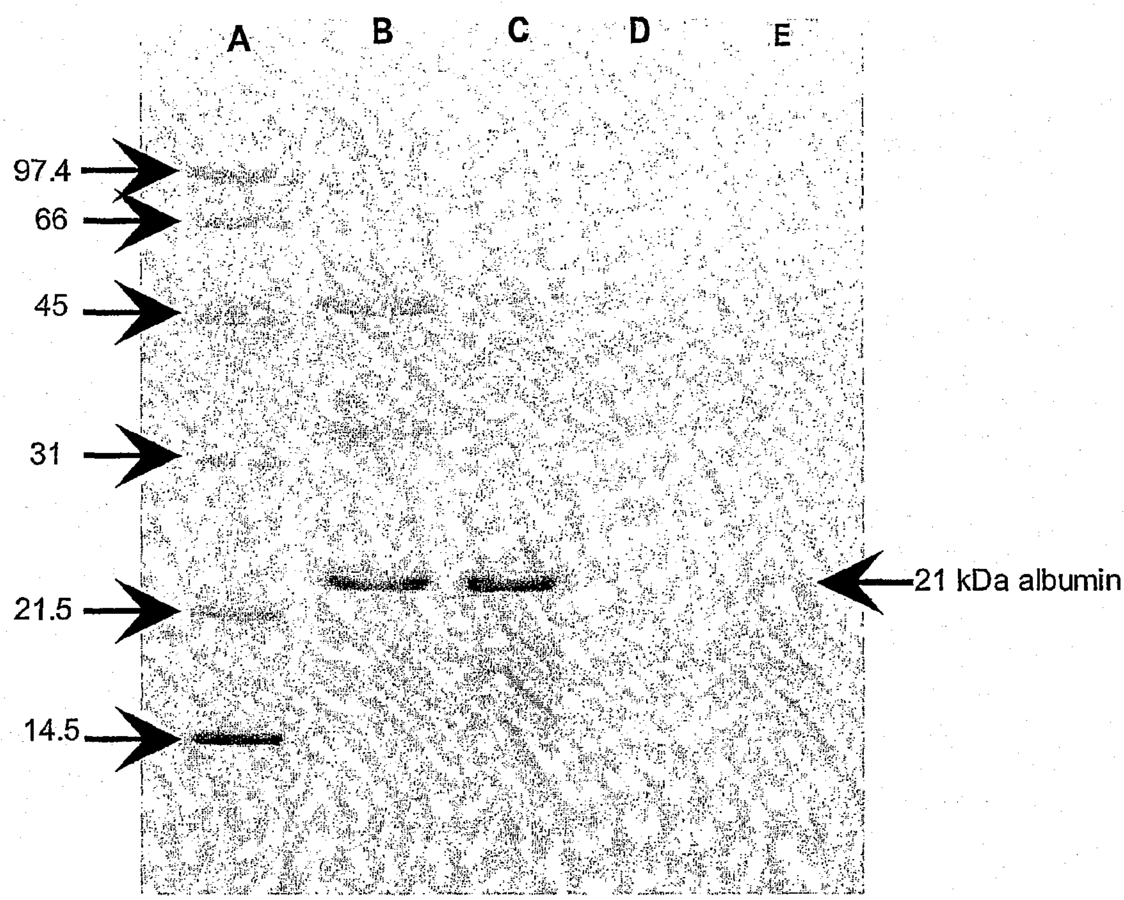


Fig. 1: SDS PAGE analysis of different extracts of cocoa acetone powder (CAP) for removal of the 21 kDa albumin.

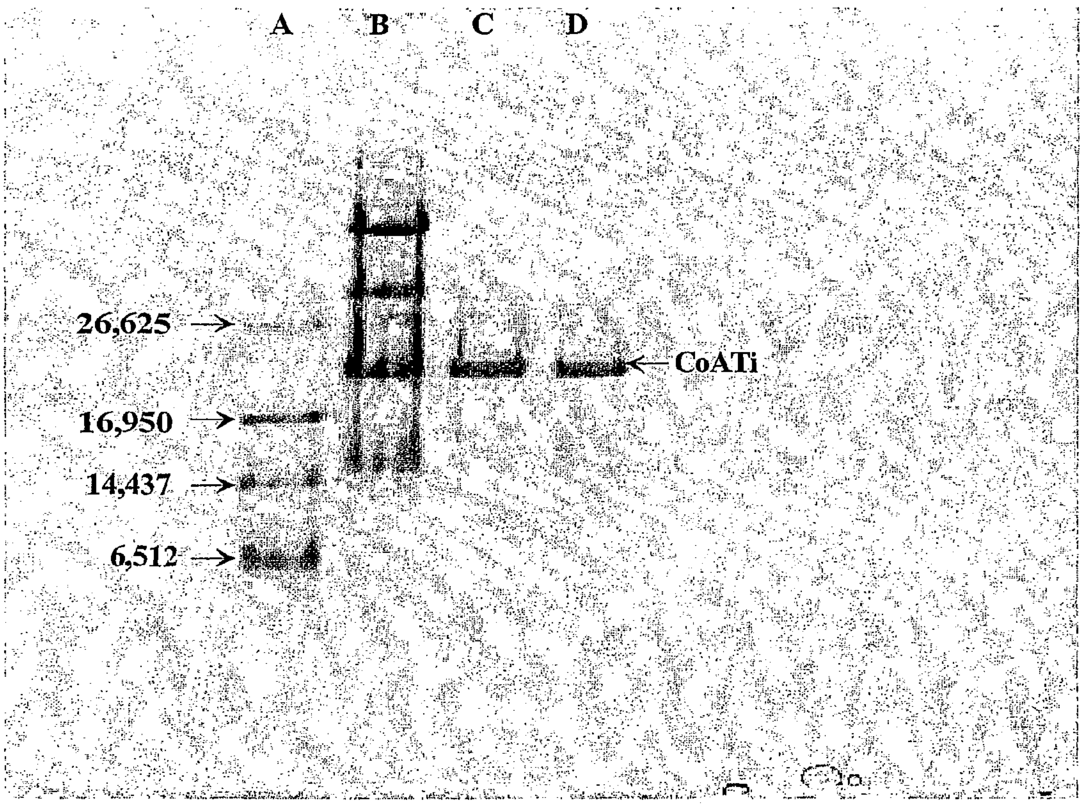


Fig. 2: Tricine-SDS-PAGE analysis of the purified 2S albumin.

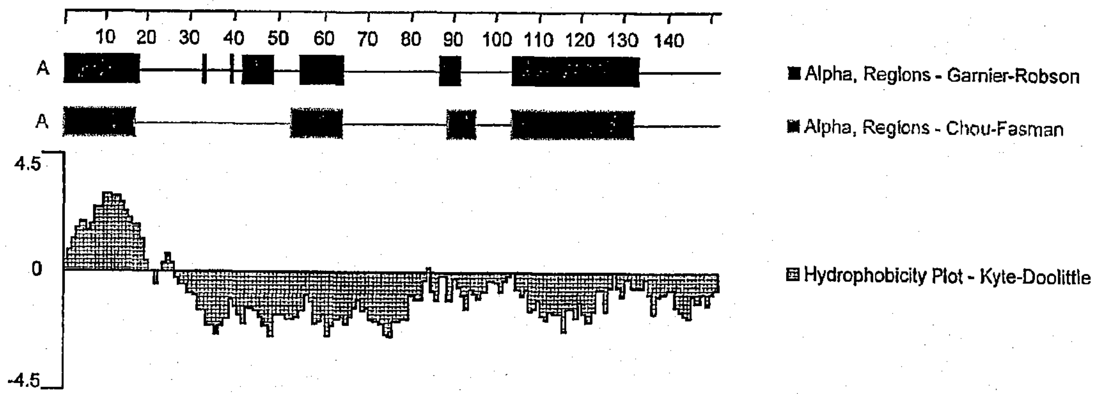
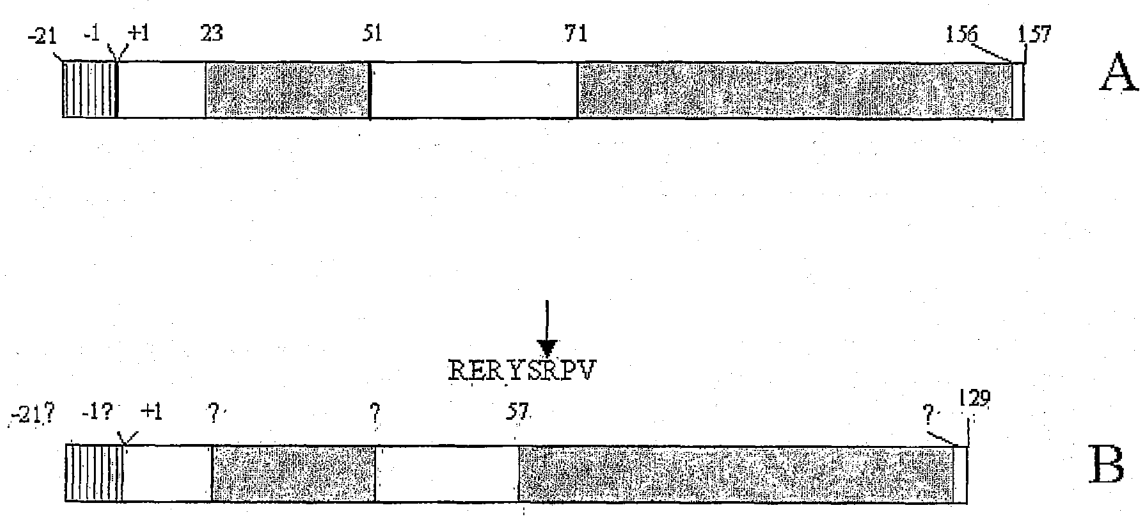


Fig. 3: Predicted α-helical regions and the hydrophobicity plot for the *T. cacao* 2S precursor protein.




```
1     RPVSKHLDSC CQQLEKLDTP CRCPGLKQAV QQQAEEGEFG REELQEMYET
51    VDKIMNKCDV EPGRCNLQPR NWF (SEQ ID No 4)
```

Fig. 4: Schematic representation of Brassica 2S protein processing (A) and the predicted processing sites for the 2S protein of *T. cacao* (B) as well as the amino acid sequence of the mature polypeptide.

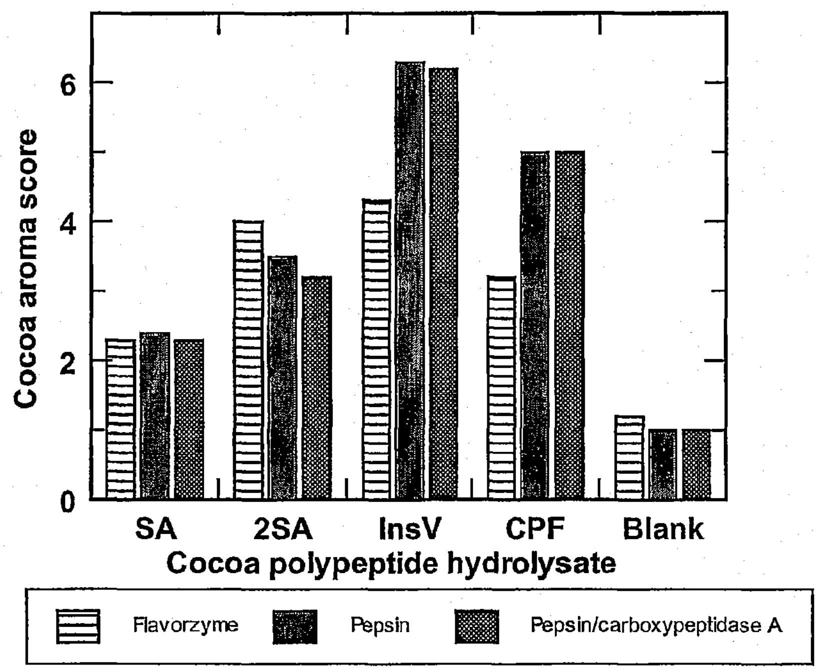


Fig. 5: Cocoa flavour evaluation of enzymatically hydrolyzed cocoa polypeptides.

SEQUENCE LISTING

<110> Societé des Produits Nestlé

<120> 2S-polypeptide of cacao

<130> 80272

<140>
<141>

<160> 5

<170> PatentIn Ver. 2.1

<210> 1
<211> 150
<212> PRT
<213> Theobroma cacao

<400> 1
Met Ala Lys Leu Gly Leu Leu Leu Ala Thr Leu Ala Leu Val Leu Phe
1 5 10 15

Leu Gly Asn Ala Ser Val Tyr His Thr Thr Val Thr Val Asp Ser Glu
20 25 30

Glu Asn Pro Trp Gly Ser Lys Glu Ser Ser Cys Gln Lys Gln Ile Lys
35 40 45

Lys Gln Asn Tyr Leu Lys His Cys Gln Glu Tyr Met Glu Glu Gln Ser
50 55 60

Arg Gly Ser Gly Ser Ser Ser Ser Arg Glu Arg Tyr Ser Arg Pro Val
65 70 75 80

Ser Lys His Leu Asp Ser Cys Cys Gln Gln Leu Glu Lys Leu Asp Thr
85 90 95

Pro Cys Arg Cys Pro Gly Leu Lys Gln Ala Val Gln Gln Gln Ala Glu
100 105 110

Glu Gly Glu Phe Gly Arg Glu Glu Leu Gln Glu Met Tyr Glu Thr Val
115 120 125

Asp Lys Ile Met Asn Lys Cys Asp Val Glu Pro Gly Arg Cys Asn Leu
130 135 140

Gln Pro Arg Asn Trp Phe
145 150

<210> 2
<211> 34
<212> PRT
<213> Theobroma cacao

<400> 2
Arg Pro Val Ser Lys His Leu Asp Ser Cys Cys Gln Gln Leu Glu Lys
1 5 10 15

Leu Asp Thr Pro Pro Arg Arg Pro Gly Leu Lys Gln Ala Val Gln Gln
20 25 30

Cys Ala

<210> 3
<211> 9
<212> PRT
<213> Theobroma cacao

<400> 3
Ser Lys Glu Xaa Ser Cys Lys Xaa Ile
1 5

<210> 4
<211> 73
<212> PRT
<213> Theobroma cacao

<400> 4
Arg Pro Val Ser Lys His Leu Asp Ser Cys Cys Gln Gln Leu Glu Lys
1 5 10 15

Leu Asp Thr Pro Cys Arg Cys Pro Gly Leu Lys Gln Ala Val Gln Gln
20 25 30

Gln Ala Glu Glu Gly Glu Phe Gly Arg Glu Glu Leu Gln Glu Met Tyr
35 40 45

Glu Thr Val Asp Lys Ile Met Asn Lys Cys Asp Val Glu Pro Gly Arg
50 55 60

Cys Asn Leu Gln Pro Arg Asn Trp Phe
65 70

<210> 5
<211> 718
<212> DNA
<213> Theobroma cacao

<400> 5
aagcagtggt aacaacgcag agtacgcggg gaagaaccaa agccttgtca tctaactagc 60
tatatatcta tatccaccat ggcaaagctc ggtctcctcc tagccaccct tgctcttgtt 120
ctcttcctcg gcaatgcctc cgtttaccac accaccgtca cggttgacag cgaggaaaac 180
ccttggggaa gcaaagagag cagctgtcag aagcagataa agaagcaaaa ctacctcaag 240
cactgtcagg agtacatgga ggagcagtcc agaggcagcg gcagcagcag cagccgtgag 300
cgctacagcc gcccсgtgag caagcaccta gactcctgtt gccagcaact ggagaagctc 360
gatacgccgt gccgttgccc tggtctaaaa caggcagtgc agcaacaggc ggaagaggga 420
gagtttggga gggaagagtt gcaagagatg tatgagacgg ttgacaagat catgaacaag 480
tgtgacgtag agcctggaag gtgtaacttg caacctcgca actggttcta gagagaaaga 540
aagatcagag ctgcctgatc taatgtaaaa caatgactgt aatgtttcac ccatcaactc 600
tggtgttcta actggaggtt tttggggtga ctagaagtac tgataatcca taaataaaag 660
cacattctcg tgtgcggttg ctttttgctt caggccaaaa aaaaaaaaaa aaaaaaaa 718

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/042327 A3

(51) International Patent Classification⁷: C12N 15/29, C07K 14/415, A23L 1/234, A61K 38/16

(21) International Application Number: PCT/EP01/13536

(22) International Filing Date: 21 November 2001 (21.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00125523.1 21 November 2000 (21.11.2000) EP

(71) Applicant (*for all designated States except US*): **SOCIETE DES PRODUITS NESTLE S.A.** [CH/CH]; P.O. Box 353, CH-1800 Vevey (CH).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **KOCHHAR, Sunil** [IN/CH]; Chemin du Grammont 2, CH-1073 Savigny (CH). **HANSEN, Carl, Erik** [NO/CH]; Chemin du Bois-Murat 20, CH-1066 Epalinges (CH). **JUILLERAT, Marcel, Alexandre** [CH/CH]; Chemin Pra-Dom-Nicod 10, Vers-Che-Les-Blancs, CH-1000 Lausanne 26 (CH). **Mc CARTHY, James** [US/FR]; 298, Vallée de Vautruchot, F-37210 Noizay (FR).

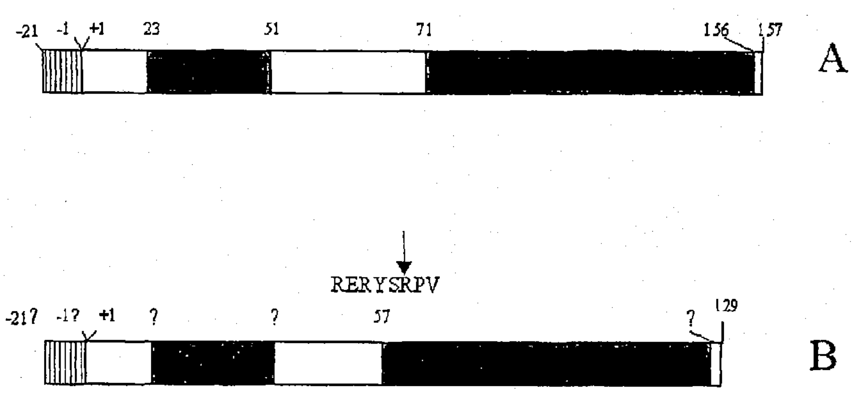
(74) Agent: **STRAUS, Alexander**; Becker, Kurig, Straus, Bavariastrasse 7, 80336 Munich (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent

[Continued on next page]

(54) Title: COCOA ALBUMIN AND ITS USE IN THE PRODUCTION OF COCOA AND CHOCOLATE FLAVOUR




```
1     RPVSKHLDSC CQQLEKLDTP CRCPGLKQAV QQQAEEGEFG REELQEMYET
51    VDKIMNKCDV EPGRCNLQPR NWF (SEQ ID No 4)
```

Schematic representation of Brassica 2S protein processing (A) and the predicted processing sites for the 2S protein of *T. cacao* (B) as well as the amino acid sequence of the mature polypeptide.

WO 02/042327 A3

(57) Abstract: A novel 2S cocoa albumin was isolated, purified and identified from cocoa beans. Enymatic hydrolysis of the protein generated a pool of flavour precursors, peptides and amin o acids that resultd in formation fo cocoa flavour upon heating with sugars. The DNA encoding a precursor cocoa 2S protein was isoltated from immature Theobroma cocoa seeds.

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 5 December 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 01/13536

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/29 C07K14/415 A23L1/234 A61K38/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A23L A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, SEQUENCE SEARCH, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	KOCHHAR SUNIL ET AL: "Isolation and characterization of 2S cocoa seed albumin storage polypeptide and the corresponding cDNA." JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY, vol. 49, no. 9, September 2001 (2001-09), pages 4470-4477, XP001104520 ISSN: 0021-8561 figure 1 ---	1-10

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 August 2002	06/09/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Espen, J

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	GALAU G A ET AL: "Cotton Mat5-A (C164) Gene and Mat5-D cDNAs encoding methionine-rich 2S albumin storage proteins" PLANT PHYSIOLOGY., vol. 99, 1992, pages 779-782, XP002169382 AMERICAN SOCIETY OF PLANT PHYSIOLOGISTS, ROCKVILLE, MD., US ISSN: 0032-0889 ---	
A	WO 98 45458 A (GUTTERIDGE STEVEN ;DU PONT (US)) 15 October 1998 (1998-10-15) ---	
A	YAVUZ M O ET AL: "Expression of the major bean proteins from Theobroma cacao (cocoa) in the yeasts Hansenula polymorpha and Saccharomyces cerevisiae" JOURNAL OF BIOTECHNOLOGY,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 46, no. 1, 18 April 1996 (1996-04-18), pages 43-54, XP004036811 ISSN: 0168-1656 -----	

INTERNATIONAL SEARCH REPORT

Information on patent family members

I tional Application No
PCT/EP 01/13536

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9845458 A	15-10-1998	AU 6883298 A	30-10-1998
		WO 9845458 A1	15-10-1998
		ZA 9801569 A	25-08-1999

Form PCT/ISA/210 (patent family annex) (July 1992)